Jeff's View
Me and my genome

I had hoped that the sequence of the human genome would tell me things about myself I had always wanted to know. But when the 'complete' sequence finally hit the front pages last year, the brouhaha was mostly about new drugs, prescient diagnostics, and higher stock prices. Few public comments were lost on the fact that this sequence was also a philosophical document.

After having made my way through the weighty issues of *Nature* and *Science* that described the draft sequence, my first reaction was disappointment. I had not thought that this 3.2 gigabyte message would say so little. I had expected 100 000 genes, and now I was supposed to have a mere 30 000–40 000, only about 10–20 times more than bacteria! I had always liked the fact that bacteria had about 1000 times less DNA than I. That felt about right. But a factor of 10–20 was carrying democracy too far.

Perhaps even this factor could explain the leap from bacteria to humans if the genes themselves were not as important as the interactions among them. Boost the number of genes 10-fold, and the possible combinations go through the roof. Such reasoning might also take care of the disquieting fact that fruit flies and worms have almost half as many genes as we do. Perhaps so, but the answer cannot be the number of genes. It must be the number of proteins. And if it comes to the number of different proteins, we leave bacteria in the dust. In fact, the dust is so dense that we cannot even guess how far ahead we are.

The simplest living cell whose genome sequence we know is *Mycoplasma genitalium*. This creature has only 580 070 base pairs of DNA and must get by with only about 470 protein-coding genes. The proteins resemble the invitees to a very exclusive party. Only the most essential players are invited – enzymes replicating, transcribing and translating genetic information, a few chaperones for protein folding, a lot of plasma membrane pumps for ions and nutrients, and a survival kit for making ATP. One or two enzymes of amino acid metabolism have crashed the party, but apart from that there is none of the usual hoi polloi. Like most exclusive parties, this one is a bore: *M. genitalium* is condemned to stay attached to more complicated cells because it must parasitize them for essential nutrients. Even so, it needs all of its exclusive proteins just to survive. Except for rare mutants, all cells of a population are therefore exactly the same. There is no biochemical room to move. If you happen to be *M. genitalium*, forget about individuality.

Free-living bacteria such as *Haemophilus influenzae* or *Escherichia coli* can breathe a little easier. According to the latest census, *H. influenzae* has about 1700 protein-coding genes, and *E. coli* K12 has about 4300. Because these bacteria can modify finished protein chains by clipping off pieces, or by attaching sugar, lipid, methyl, or phosphoryl groups, the actual number of different proteins probably exceeds the number of protein-coding genes. We do not know by how much, but the difference may be less than twofold. Despite the fact that these free-living bacteria have about 10 times as many proteins as *M. genitalium*, they still need most of them to survive in the wild. In most respects, all cells of a population (again discounting rare mutants) are therefore identical. But not always. These bacteria usually have rotating flagella coupled to chemosensors by which they can swim towards food, or away from poison. The molecular principle underlying this simple yes/no decision is surprisingly similar to that governing the more complex decisions in our own brain. And these decisions are not always predictable. Look at a swarm of *E. coli* under the light microscope and add a drop of glucose solution to one edge of the cover slide: some cells will immediately start to swim straight at the food, whereas others will have trouble making up their mind, or keep a straight course. The cells have the same genes and the same environment, yet behave differently. When they look for food, they show some individuality. Not much to write home about, but still impressive for cells with only a few million base pairs' worth of DNA.

With its 3300 million base pairs, our own genome is much larger, and immensely more mysterious. Less than 1.5% represents typical genes, and we have no idea what the rest is good for. But we do know quite a bit about how we read these genes. That's the department in which we really shine. We can, of course, read them from the beginning to the end just like bacteria do, but we may also start later, finish earlier, or skip sections in between. We can play similar tricks even when translating messenger RNA. Yet our ingenuity really takes off once we have finished a polypeptide chain. We can cut away pieces from either end with proteases, or attach an astonishing assortment of chemical groups that may affect the protein's function, its intracellular location, its half-life, or its association with other proteins. Our cells have at least 1000 protein kinases whose major, if not only job it is to hook a phosphoryl group onto another protein. The magic wand of posttranslational modification gives our proteome polychromatic glitter. To add to this glitter, almost each gene in our body cells exists in two copies that may differ from each other. There may also be thousands or even tens of thousands of very small proteins that we are not even aware of. They are so small that they run off our usual SDS–polyacrylamide gels, and that their open reading frames are invisible to our search algorithms. At yet another level, each cell has dozens, hundreds or even thousands of mitochondrial genomes that are not always exactly identical and whose protein products may interact in still unknown ways with the proteins encoded by nuclear genes. By reshuffling and hyper-mutating some of their genes, the cells of our immune system can theoretically make a quasi-infinite number of different immune proteins. And to cap it all, there is evidence to suggest that mammalian brain cells can alter the amount or the properties of some of their neurological switch proteins in response to training or other external stimuli. How many different proteins can we make? It is anybody's guess. A conservative estimate would be around 100 000. My personal bet would be closer to half a million. And that's not counting the immune proteins.

As impressive as the protein spectrum of our cells is, the real marvel is its regulation. Higher eukaryotes have im-

0014-5793/01/$20.00 © 2001 Published by Elsevier Science B.V. on behalf of the Federation of European Biochemical Societies.
PII: S0014-5793(01)02876-9

mensely complex devices for fine-tuning the expression of their genes. Some of these devices bind to regulatory gene sequences and can act quickly, whereas others alter the long-range structure of DNA and can shut off a gene for a lifetime. Most of these devices have protein subunits that sense what needs to be done. If these devices go out of control, disaster follows. The versatility, subtlety, and mind-boggling complexity of our gene regulation far exceed anything in bacteria. The intricate tapestry of our proteome changes constantly, and we are a long way from understanding how changing one component ripples through the entire system.

Bacteria read their genome. We interpret ours. Our genome is not a pedantically annotated score that leaves the conductor little freedom, but a general base from which we can evoke many different types of music. Our genome is so rich because we can read it in so many different ways.

We can make so many diverse proteomes that each of us is unique. This even holds for identical twins: Boris Becker's hypothetical identical twin would probably look like his famous brother, but might well be an average tennis player. The immense information hidden within our genome is the grace that grants each of us individuality.

There is no tyrant as merciless as the small genome. It allows no biological freedom and forbids individuality. The more information a genome carries, the greater is its magnanimity in allowing different phenotypes. To me, the information content of a genome ranks an organism in the hierarchy of life. If organisms have biological dignity, then this dignity must be related to genomic information.

Yet a chimpanzee, or even a mouse, has as much DNA and about as many genes as I do. And it is highly unlikely that I owe my humanity to a few key genes. Counting base pairs or genes may be good enough to sketch the tree of life, but not nearly good enough to delineate the ramification of its topmost branches. Some essential feature of my genome still escapes me. I still do not know why I have large frontal lobes, walk upright, and love to play with words. If I want to tell a chimp his place, I cannot (yet) flaunt my genome.

On the other hand, this genome tells me much about where I come from. The distribution of point mutations in linked genes from people round the world suggests that I am the offspring of a very small group of humans that split off from a much older African population between 27 000 and 53 000 years ago. Just think of it – all of us Northern Europeans come from a few hundred individuals! Was this biological bottleneck caused by the Ice Age, or by a devastating disease? It sure was a close call.

Our body as a molecular system is so complex that we cannot quantitatively predict its behavior. Perhaps we shall never be able to do this, because a system that complex may defy rigorous prediction. Our cells might have so many parts that they behave stochastically. How wonderful! At long last we would know that we are not merely biochemical machines run by a fixed set of genes. The complexity of our cells would release us from the prison of determinism. It has been argued that a modern airplane with its several hundred thousand parts is as complex as living cells, yet does not behave stochastically. This is true, even though many frequent flyers may have their doubts. But the comparison is not fair. The parts of an airplane are invariable, whereas those of a cell fluctuate constantly. Airplanes that automatically expand, shrink, or even jettison their parts depending on the flying conditions might indeed be unpredictable, yet safer than present ones.

I am no longer disappointed by the fact that my genome has so few genes. At least it is not a tyrant. I had been afraid that it would be a book of laws. It feels good that it is just a set of rules.

I thank Stuart J. Edelstein, Susan Gasser and Ueli Schibler for helpful comments.

Gottfried Schatz
Swiss Science and Technology Council,
Bern, Switzerland
E-mail address: Gottfried.Schatz@unibas.ch

Note from the Editorial Office:

'*Jeff's View*' is the first in a series of invited commentaries from Gottfried Schatz, a distinguished scientist who presently heads the Swiss Science and Technology Council. These articles will deal with novel developments in biochemistry and molecular cell biology as well as general research and policy issues in the field of life science. We look forward to receiving more such lively and colorful contributions from Jeff during the upcoming year. Any conclusions or opinions expressed in these commentaries, however, are not necessarily the views of the Federation of European Biochemical Societies (FEBS) or of *FEBS Letters*.